

Entropy Jumps for Radially Symmetric Random Vectors

Thomas A. Courtade

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

November 4, 2016

Abstract

We establish a quantitative bound on the entropy jump associated to the sum of independent, identically distributed (IID) radially symmetric random vectors having dimension greater than one. Following the usual approach, we first consider the analogous problem of Fisher information dissipation, and then integrate along the Ornstein-Uhlenbeck semigroup to obtain an entropic inequality. In a departure from previous work, we appeal to a result by Desvillettes and Villani on entropy production associated to the Landau equation. This obviates strong regularity assumptions, such as presence of a spectral gap and log-concavity of densities, but comes at the expense of radial symmetry. As an application, we give a quantitative estimate of the deficit in the Gaussian logarithmic Sobolev inequality for radially symmetric functions.

1 Introduction

Let X be a random vector on \mathbb{R}^d with density f . The entropy associated to X is defined by

$$h(X) = - \int_{\mathbb{R}^d} f \log f, \quad (1)$$

provided the integral exists. The non-Gaussianness of X , denoted by $D(X)$, is given by

$$D(X) = h(G^X) - h(X), \quad (2)$$

where G^X denotes a Gaussian random vector with the same covariance as X . Evidently, $D(X)$ is the relative entropy of X with respect to G^X , and is therefore nonnegative. Moreover, $D(X) = 0$ if and only if X is Gaussian.

Our main result may be informally stated as follows: Let X, X_* be IID radially symmetric random vectors on \mathbb{R}^d , $d \geq 2$, with sufficiently regular density f . For any $\varepsilon > 0$

$$h\left(\frac{1}{\sqrt{2}}(X + X_*)\right) - h(X) \geq C_\varepsilon(X) D(X)^{1+\varepsilon}, \quad (3)$$

where $C_\varepsilon(X)$ is an explicit function depending on ε , the regularity of f , and a finite number of moments of X . In particular, if $\mathbb{E}|X|^{4+\delta} < \infty$ for some $\delta > 0$, then $C_{2/\delta}(X) > 0$. A precise statement can be found in Section 3, along with an analogous result for Fisher information and a related estimate that imposes no regularity conditions. In interpreting the result, it is important

to note that, although a radially symmetric density $f : \mathbb{R}^d \rightarrow [0, \infty)$ has a one-dimensional parameterization, the convolution $f * f$ is inherently a d -dimensional operation unless f is Gaussian. Thus, it does not appear that (3) can be easily reduced to a one-dimensional problem.

The quantity $h(\frac{1}{\sqrt{2}}(X + X_*)) - h(X)$ characterizes the *entropy production* (or, entropy jump) associated to X under rescaled convolution. Similarly, letting $J(X) = 4 \int_{\mathbb{R}^d} |\nabla \sqrt{f}|^2$ denote the Fisher information of X , the quantity $J(X) - J(\frac{1}{\sqrt{2}}(X + X_*))$ characterizes the *dissipation of Fisher information*. By the convolution inequalities of Shannon [1], Blachman [2] and Stam [3] for entropy power and Fisher information, it follows that both the production of entropy and the dissipation of Fisher information under rescaled convolution are nonnegative. Moreover, both quantities are identically zero if and only if X is Gaussian.

The fundamental problem of bounding entropy production (and dissipation of Fisher information) has received considerable attention, yet quantitative bounds are few. In particular, the entropy power inequality establishes that entropy production is strictly greater than zero unless X is Gaussian, but gives no indication of how entropy production behaves as, say, a function of $D(X)$ when X is non-Gaussian. As a consequence, basic stability properties of the entropy power inequality remain elusive, despite it being a central inequality in information theory. For instance, a satisfactory answer to the following question is still out of reach: *If the entropy power inequality is nearly saturated, are the random summands involved quantifiably close to Gaussian?*

Perhaps the first major result to address the question of entropy production in this setting is due to Carlen and Soffer [4], who showed for each random vector X with $J(X) < \infty$ and $\text{Cov}(X) = I$, there exists a nonnegative function $\Theta_{\psi, \chi}$ on $[0, \infty)$, strictly increasing from 0, and depending only on the two auxiliary functions

$$\psi(R) = \mathbb{E}[|X|^2 \mathbf{1}_{\{|x| \geq R\}}], \quad (4)$$

and

$$\chi(t) = h(X_t) - h(X), \quad X_t := e^{-t}X + (1 - e^{-2t})^{1/2}G^X, \quad (5)$$

such that

$$h(\frac{1}{\sqrt{2}}(X + X_*)) - h(X) \geq \Theta_{\psi, \chi}(D(X)). \quad (6)$$

Moreover, the function Θ_{ψ_0, χ_0} will do for any density f with $\psi_f \leq \psi_0$ and $\chi_f \leq \chi_0$. Hence, this provides a nonlinear estimate of entropy production in terms of $D(X)$ that holds uniformly over all probability densities that exhibit the same decay and smoothness properties (appropriately defined). Unfortunately, the proof establishing existence of $\Theta_{\psi, \chi}$ relies on a compactness argument, and therefore falls short of giving satisfactory quantitative bounds.

A random vector X with density f has spectral gap $c > 0$ (equivalently, finite Poincaré constant $1/c$) if, for all smooth functions g with $\int_{\mathbb{R}^d} f g = 0$,

$$c \int_{\mathbb{R}^d} f g^2 \leq \int_{\mathbb{R}^d} f |\nabla g|^2. \quad (7)$$

Generally speaking, a non-zero spectral gap is a very strong regularity condition on the density f (e.g., it implies X has finite moments of all orders).

In dimension $d = 1$, if X has spectral gap $c > 0$, then [5, 6] established the linear bound¹

$$h\left(\frac{1}{\sqrt{2}}(X + X_*)\right) - h(X) \geq \frac{c}{2+2c}D(X). \quad (8)$$

In dimension $d \geq 2$, Ball and Nguyen [8] recently established an essentially identical result under the additional assumption that X is isotropic (i.e., $\text{Cov}(X) = \text{I}$) with log-concave density f . Along these lines, Toscani has a strengthened EPI for log-concave densities [9], but the deficit is qualitative in nature in contrast to the quantitative estimate obtained by Ball and Nguyen.

Clearly, entropy production and Fisher information dissipation is closely related to convergence rates in the entropic and Fisher information central limit theorems. Generally speaking though, bounds in the spirit of (8) are unnecessarily strong for establishing entropic central limit theorems of the form $D(S_n) = O(1/n)$, where $S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ denotes the normalized sum of n IID copies of X . Indeed, it was long conjectured that $D(S_n) = O(1/n)$ under moment conditions. This was positively resolved by Bobkov, Chistyakov and Götze [10, 11] using Edgeworth expansions and local limit theorems. By Pinsker's inequality, we know that $D(X)$ dominates squared total-variation distance, so $D(S_n) = O(1/n)$ is interpreted as a version of the Berry-Esseen theorem for the entropic CLT with the optimal convergence rate. However, while the results of Bobkov et al. give good long-range estimates of the form $D(S_n) = O(1/n)$ (with explicit constants depending on the cumulants of X), the smaller-order terms propagate from local limit theorems for Edgeworth expansions and are non-explicit. Thus, explicit bounds for the initial entropy jump $h(\frac{1}{\sqrt{2}}(X + X_*)) - h(X)$ cannot be readily obtained.

Along these lines, we remark that Ledoux, Nourdin and Peccati [12] recently established the weaker convergence rate $D(S_n) = O(\frac{\log n}{n})$ via the explicit bound

$$D(S_n) \leq \frac{S^2(X|G)}{2n} \log \left(1 + n \frac{I(X|G)}{S^2(X|G)} \right), \quad (9)$$

where $I(X|G)$ is the relative Fisher information of X with respect to the standard normal G , and $S(X|G)$ denotes the *Stein discrepancy* of X with respect to G , which is defined when a Stein kernel exists (see [12] for definitions). In principle, this has potential to give an explicit bound on the entropy production $h(\frac{1}{\sqrt{2}}(X + X_*)) - h(X)$ in terms of $S^2(X|G)$ by considering $n = 2$. Unfortunately, a Stein kernel may not always exist; even if it does, further relationships between $S^2(X|G)$ and $I(X|G)$ or $D(X)$ would need to be developed to ultimately yield a bound like (3).

Another related line of work in statistical physics considers quantitative bounds on entropy production in the Boltzmann equation (see the review [13] for an overview). The two problems are not the same, but there is a strong analogy between entropy production in the Boltzmann equation and entropy jumps associated to rescaled convolution as can be seen by comparing [14] to [4]. The details of this rich subject are tangential to the present discussion, but we remark that a major milestone in this area was achieved when the entropy production in the Boltzmann equation was bounded from below by an explicit function of $D(X)$ (and various norms of X), where X models the velocity of a particle in a rarified gas [15, 16]. A key ingredient used to prove this bound was an earlier result by Desvillettes and Villani that controls relative Fisher information $I(X|G)$ via entropy production in the Landau equation:

¹Technically speaking, Barron and Johnson establish a slightly different inequality. However, a modification of their argument gives the same result as Ball, Barthe and Naor. See the discussion surrounding [7, Theorem 2.4].

Lemma 1. [17] Let X be a random vector on \mathbb{R}^d , satisfying $\mathbb{E}|X|^2 = d \geq 2$ and having density f . Then,

$$\frac{1}{2} \iint |x - x_*|^2 f(x) f(x_*) \left| \Pi(x - x_*) \left[\frac{\nabla f}{f}(x) - \frac{\nabla f}{f}(x_*) \right] \right|^2 dx dx_* \geq \lambda(d-1)I(X|G), \quad (10)$$

where λ is the minimum eigenvalue of the covariance matrix associated to X , and $\Pi(v)$ is the orthogonal projection onto the subspace orthogonal to $v \in \mathbb{R}^d$.

Our proof of (3) follows a program similar to [15, 16], and is conceptually straightforward after the correct ingredients are assembled. In particular, we begin by recognizing that the LHS of (10) resembles dissipation of Fisher information when written in the context of L^2 projections (cf. [6, Lemma 3.1]). Using the radial symmetry assumption, we are able to bound the Fisher information dissipation from below by error terms plus entropy production in the Landau equation, which is subsequently bounded by relative Fisher information using Lemma 1. Care must be exercised in order to control error terms (this is where our regularity assumptions enter), but the final result (3) closely parallels that proved in [15] for the Boltzmann equation. We remark that the assumption of a non-vanishing Boltzmann collision kernel in [15] has a symmetrizing effect on the particle density functions involved; the rough analog in the present paper is the radial symmetry assumption.

Organization

The rest of this paper is organized as follows. Section 2 briefly introduces notation and definitions that are used throughout. Main results are stated and proved in Section 3, followed by a brief discussion on potential extensions to non-symmetric distributions. Section 4 gives an application of the results to bounding the deficit in the Gaussian logarithmic Sobolev inequality.

2 Notation and Definitions

For a vector $v \in \mathbb{R}^d$, we let $|v| := (\sum_{i=1}^d v_i^2)^{1/2}$ denote its Euclidean norm. For a random variable X on \mathbb{R} and $p \geq 1$, we write $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$ for the usual L^p -norm of X . It will be convenient to use the same notation for $0 < p < 1$, with the understanding that $\|\cdot\|_p$ is not a norm in this case.

Throughout, $G \sim N(0, I)$ denotes a standard Gaussian random vector on \mathbb{R}^d ; the dimension will be clear from context. For a random vector X on \mathbb{R}^d , we let $G^X = \sqrt{d^{-1}\mathbb{E}|X|^2}G$ be a normalized Gaussian vector, so that $\mathbb{E}|X|^2 = \mathbb{E}|G^X|^2$. For $d \geq 2$, we denote the coordinates of a random vector X on \mathbb{R}^d as (X_1, X_2, \dots, X_d) . Thus, for example, G_1^X is a zero-mean Gaussian random variable with variance $d^{-1}\mathbb{E}|X|^2$.

For a random vector X with smooth density² f , we define the Fisher information

$$J(X) = 4 \int \left| \nabla \sqrt{f} \right|^2 = \int_{f>0} \frac{|\nabla f|^2}{f} \quad (11)$$

and the entropy

$$h(X) = - \int f \log f, \quad (12)$$

²All densities are with respect to Lebesgue measure.

where ‘log’ denotes the natural logarithm throughout. For random vectors X, Q with respective densities f, g , the relative Fisher information is defined by

$$I(X|Q) = 4 \int g \left| \nabla \sqrt{f/g} \right|^2 \quad (13)$$

and the relative entropy is defined by

$$D(X|Q) = \int f \log \frac{f}{g}. \quad (14)$$

Evidently, both quantities are nonnegative and

$$I(X) := I(X|G^X) = J(X) - J(G^X) \quad D(X) := D(X|G^X) = h(G^X) - h(X). \quad (15)$$

Finally, we recall two basic inequalities that will be taken for granted several times without explicit reference: for real-valued a, b we have $(a+b)^2 \leq 2a^2 + 2b^2$, and for random variables X, Y , we have Minkowski’s inequality: $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$ when $p \geq 1$.

Definition 1. A random vector X with density f is radially symmetric if $f(x) = \phi(|x|)$ for some function $\phi : \mathbb{R} \rightarrow [0, \infty)$.

We primarily concern ourselves with random vectors that satisfy certain mild regularity conditions. In particular, it is sufficient to control $|\nabla \log f(x)|$ pointwise in terms of $|x|$.

Definition 2. A random vector X on \mathbb{R}^d with smooth density f is c -regular if, for all $x \in \mathbb{R}^d$,

$$|\nabla \log f(x)| \leq c(|x| + \mathbb{E}|X|). \quad (16)$$

We remark that the smoothness requirement of f in the definition of c -regularity is stronger than generally required for our purposes. However, it allows us to avoid further qualifications; for instance, the identities (11) hold for any c -regular function. Moreover, since $\nabla \log f = \frac{\nabla f}{f}$ for smooth f , we have $J(X) < \infty$ for any c -regular X with $\mathbb{E}|X|^2 < \infty$.

Evidently, c -regularity quantifies the smoothness of a density function. The following important example shows that any density can be mollified to make it c -regular.

Proposition 1. [18] Let X and Z be independent, where $Z \sim N(0, \sigma^2 \mathbf{I})$. Then $Y = X + Z$ is c -regular with $c = 4/\sigma^2$.

Observe that, in the notation of the above proposition, if X is radially symmetric then so is Y . Therefore, Proposition 1 provides a convenient means to construct radially symmetric random vectors that are c -regular. Indeed, we have the following useful corollaries (proofs are found in the appendix).

Proposition 2. Let X be a random vector on \mathbb{R}^d , and let $X_t = e^{-t}X + (1 - e^{-2t})^{1/2}G$ for $t \geq 0$. If X is c -regular, then X_t is $(5ce^{2t})$ -regular.

Proposition 3. Let R_0 be a non-negative random variable with $\mathbb{E}R_0^2 = 1$ and distribution function F_{R_0} . For any $d \geq 2$ and $t, \varepsilon > 0$, there exists a $(4/\varepsilon)$ -regular radially symmetric random vector X on \mathbb{R}^d with $\mathbb{E}|X|^2 = d$ and $R = \frac{1}{\sqrt{d}}|X|$ satisfying

$$F_{R_0} \left(\frac{r - \sqrt{(t+1)\varepsilon}}{\sqrt{1-\varepsilon}} \right) - e^{-dt^2/8} \leq F_R(r) \leq F_{R_0} \left(\frac{r + \sqrt{(t+1)\varepsilon}}{\sqrt{1-\varepsilon}} \right) + e^{-dt^2/8}, \quad (17)$$

where F_R is the distribution function of R .

3 Main Results

In this section, we establish quantitative estimates on entropy production and Fisher information dissipation under rescaled convolution. As can be expected, we begin with an inequality for Fisher information, and then obtain a corresponding entropy jump inequality by integrating along the Ornstein-Uhlenbeck semigroup.

3.1 Dissipation of Fisher Information under Rescaled Convolution

Theorem 1. *Let X, X_* be IID radially symmetric random vectors on \mathbb{R}^d , $d \geq 2$, with c -regular density f . For any $\varepsilon > 0$*

$$J(X) - J\left(\frac{1}{\sqrt{2}}(X + X_*)\right) \geq K_\varepsilon(X)I(X)^{1+\varepsilon}, \quad (18)$$

where

$$K_\varepsilon(X) = \frac{(\varepsilon/8)^\varepsilon}{(8(1+\varepsilon))^{1+\varepsilon}} \cdot \frac{\| |X|^2 \|_1^{1+\varepsilon}}{c^{2\varepsilon} \| |X|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}}. \quad (19)$$

Remark 1. *We have made no attempt to optimize the constant $K_\varepsilon(X)$.*

A few comments are in order. First, we note that inequality (18) is invariant to scaling $t : X \mapsto tX$ for $t > 0$. Indeed, if X is c -regular, then a change of variables shows that tX is (c/t^2) -regular. So, using homogeneity of the norms, we find that

$$K_\varepsilon(tX) = t^{2\varepsilon} K_\varepsilon(X). \quad (20)$$

Combined with the property that $t^2 J(tX) = J(X)$, we have

$$K_\varepsilon(tX)I(tX|G^{tX})^{1+\varepsilon} = t^{-2} K_\varepsilon(X)I(X|G^X)^{1+\varepsilon}, \quad (21)$$

which has the same scaling behavior as the LHS of (18). That is,

$$J(tX) - J\left(\frac{1}{\sqrt{2}}(tX + tX_*)\right) = t^{-2} \left(J(X) - J\left(\frac{1}{\sqrt{2}}(X + X_*)\right) \right). \quad (22)$$

Second, inequality (18) does not contain any terms that explicitly depend on dimension. However, it is impossible to say that inequality (18) is dimension-free in the usual sense that both sides scale linearly in dimension when considering product distributions. Indeed, the product of two identical radially symmetric densities is again radially symmetric if and only if the original densities were Gaussian themselves, which corresponds to the degenerate case when the dissipation of Fisher information is identically zero. However, inequality (18) does exhibit dimension-free behavior in the following sense: Suppose for simplicity that X is normalized so that $\mathbb{E}|X|^2 = d$. Since X is radially symmetric, it can be expressed as the product of independent random variables $X = \sqrt{d}RU$, where U is uniform on the $(d-1)$ -dimensional sphere \mathbb{S}^{d-1} and $R = \frac{1}{\sqrt{d}}|X|$ is a nonnegative real-valued random variable satisfying $\mathbb{E}R^2 = 1$. Now, by the log Sobolev inequality and Talagrand's inequality, we have

$$I(X|G) \geq 2D(X|G) \geq W_2^2\left((\sqrt{d}RU), G\right) = dW_2^2\left(R, \frac{1}{\sqrt{d}}|G|\right) \quad (23)$$

The equality follows since, for any vectors v, w we have $|v - w|^2 \geq ||v| - |w||^2$. However, this can be achieved with equality by the coupling $G = |G|U$. Thus, we have

$$K_\varepsilon(X)I(X|G)^{1+\varepsilon} \geq \frac{(\varepsilon/8)^\varepsilon d^{1+2\varepsilon} W_2^{2\varepsilon}\left(R, \frac{1}{\sqrt{d}}|G|\right)}{c^{2\varepsilon} (8(1+\varepsilon))^{1+\varepsilon} \|X\|^2_{2+1/\varepsilon}^{1+2\varepsilon}} I(X|G) \quad (24)$$

$$= \frac{(\varepsilon/8)^\varepsilon W_2^{2\varepsilon}\left(R, \frac{1}{\sqrt{d}}|G|\right)}{c^{2\varepsilon} (8(1+\varepsilon))^{1+\varepsilon} \|R\|_{2+1/\varepsilon}^{1+2\varepsilon}} I(X|G). \quad (25)$$

Now, we note that $\mathbb{E}R^2 = \mathbb{E}(\frac{1}{\sqrt{d}}|G|)^2 = 1$, so we have a bound of the form

$$J(X) - J(\frac{1}{\sqrt{2}}(X + X_*)) \geq \tilde{K}_\varepsilon(X)I(X|G^X), \quad (26)$$

where the function $\tilde{K}_\varepsilon(X)$ is effectively dimension-free in that it only depends on the (one-dimensional) quadratic Wasserstein distance between R and $\frac{1}{\sqrt{d}}|G|$. For $d \rightarrow \infty$, the law of large numbers implies that $\frac{1}{\sqrt{d}}|G| \rightarrow 1$ a.s. Therefore, $W_2^2\left(R, \frac{1}{\sqrt{d}}|G|\right)$ behaves similarly to $\mathbb{E}|R - 1|^2 \geq \text{Var}(R)$ high dimensions. Indeed, by the triangle inequality applied to W_2 ,

$$\left|W_2\left(R, \frac{1}{\sqrt{d}}|G|\right) - \|R - 1\|_2\right| \leq \left\|\frac{1}{\sqrt{d}}|G| - 1\right\|_2 = O\left(\frac{1}{\sqrt{d}}\right). \quad (27)$$

So, we see that (18) depends very weakly on d when the marginal distribution of $|X|$ is preserved and dimension varies.

One important question remains: As dimension $d \rightarrow \infty$, do there exist random vectors X on \mathbb{R}^d with sufficient regularity for which the associated random variable R is not necessarily concentrated around 1? The answer to this is affirmative in the sense of Proposition 3: we may approximate any distribution function F_{R_0} to within arbitrary accuracy, at the (potential) expense of increasing the regularity parameter c .

Proof of Theorem 1. As remarked above, inequality (18) is invariant to scaling. Hence, there is no loss of generality in assuming that X is normalized according to $\mathbb{E}|X|^2 = d$. Also, since X is radially symmetric, $X - X_*$ is equal to $X + X_*$ in distribution, therefore we seek to lower bound the quantity

$$J(X) - J(\frac{1}{\sqrt{2}}(X + X_*)) = J(X) - 2J(X - X_*). \quad (28)$$

Toward this end, define $W = X - X_*$, and denote its density by f_W . By the projection property of the score function of sums of independent random variables, the following identity holds (e.g., [7, Lemma 3.4]):

$$2(J(X) - 2J(X - X_*)) = \mathbb{E}|2\rho_W(W) - (\rho(X) - \rho(X_*))|^2, \quad (29)$$

where $\rho = \nabla \log f$ is the score function of X and $\rho_W = \nabla \log f_W$ is the score function of W .

For $v \in \mathbb{R}^d$, let $\Pi(v)$ denote the orthogonal projection onto the subspace orthogonal to v . Now, we have

$$2J(X) - 4J(X - X_*) = \mathbb{E}|2\rho_W(W) - (\rho(X) - \rho(X_*))|^2 \quad (30)$$

$$\geq \mathbb{E}|2\Pi(W)\rho_W(W) - \Pi(X - X_*)(\rho(X) - \rho(X_*))|^2 \quad (31)$$

$$= \mathbb{E}|\Pi(X - X_*)(\rho(X) - \rho(X_*))|^2. \quad (32)$$

The inequality follows since $\Pi(w) = \Pi(x - x_*)$ by definition, and $|v| \geq |\Pi(w)v|$ since $\Pi(w)$ is an orthogonal projection. The last equality follows since $\Pi(w)\rho_W(w) = 0$ due to the fact that $\Pi(w)\nabla f_W(w)$ is the tangential gradient of f_W , which is identically zero due to radial symmetry of f_W .

Next, for any $R > 0$, use the inequality

$$1 \geq \frac{|x - x_*|^2}{R^2} - \frac{|x - x_*|^2}{R^2} \mathbf{1}_{\{|x - x_*| > R\}} \quad (33)$$

to conclude that

$$\begin{aligned} 2J(X) - 4J(X - X_*) &\geq \frac{1}{R^2} \mathbb{E} \left[|X - X_*|^2 |\Pi(X - X_*) (\rho(X) - \rho(X_*))|^2 \right] \\ &\quad - \frac{1}{R^2} \mathbb{E} \left[|X - X_*|^2 |\Pi(X - X_*) (\rho(X) - \rho(X_*))|^2 \mathbf{1}_{\{|X - X_*| > R\}} \right]. \end{aligned} \quad (34)$$

We bound the second term first. By c -regularity and the triangle inequality, we have

$$|\Pi(x - x_*) (\rho(x) - \rho(x_*))| \leq |\rho(x) - \rho(x_*)| \leq c(|x| + |x_*|) + 2c\mathbb{E}|X|. \quad (35)$$

So, noting the inclusion

$$\{|x - x_*| > R\} \supseteq \{|x| \geq R/2, |x_*| \leq |x|\} \cup \{|x_*| \geq R/2, |x| \leq |x_*|\}, \quad (36)$$

we have the pointwise inequality

$$\mathbf{1}_{\{|x - x_*| > R\}} |x - x_*|^2 |\Pi(x - x_*) (\rho(x) - \rho(x_*))|^2 \quad (37)$$

$$\leq \mathbf{1}_{\{|x| \geq R/2, |x_*| \leq |x|\} \cup \{|x_*| \geq R/2, |x| \leq |x_*|\}} |x - x_*|^2 |\rho(x) - \rho(x_*)|^2 \quad (38)$$

$$\leq \mathbf{1}_{\{|x| \geq R/2\}} 4|x|^2 (2c|x| + 2c\mathbb{E}|X|)^2 + \mathbf{1}_{\{|x_*| \geq R/2\}} 4|x_*|^2 (2c|x_*| + 2c\mathbb{E}|X|)^2. \quad (39)$$

Taking expectations and using the fact that X, X_* are IID, we have for any conjugate exponents $p, q \geq 1$ and $\beta > 0$,

$$\begin{aligned} &\mathbb{E} \left[|X - X_*|^2 |\Pi(X - X_*) (\rho(X) - \rho(X_*))|^2 \mathbf{1}_{\{|X - X_*| > R\}} \right] \\ &\leq 16c^2 \mathbb{E} \left[|X|^2 (|X| + \mathbb{E}|X|)^2 \mathbf{1}_{\{|X| \geq R/2\}} \right] \end{aligned} \quad (40)$$

$$\leq 32c^2 \mathbb{E} \left[|X|^4 \mathbf{1}_{\{|X| > R/2\}} \right] + 32c^2 (\mathbb{E}|X|)^2 \mathbb{E} \left[|X|^2 \mathbf{1}_{\{|X| \geq R/2\}} \right] \quad (41)$$

$$\leq 32c^2 \| |X|^2 \|_{2p}^2 (\Pr \{|X| \geq R/2\})^{1/q} + 32c^2 (\mathbb{E}|X|)^2 \| |X|^2 \|_p (\Pr \{|X| \geq R/2\})^{1/q} \quad (42)$$

$$\leq 32c^2 \| |X|^2 \|_{2p}^2 \left(\frac{\mathbb{E}|X|^\beta}{(R/2)^\beta} \right)^{1/q} + 32c^2 (\mathbb{E}|X|)^2 \| |X|^2 \|_p \left(\frac{\mathbb{E}|X|^\beta}{(R/2)^\beta} \right)^{1/q} \quad (43)$$

$$= \frac{32 \cdot 2^{\beta/q} c^2}{R^{\beta/q}} \left(\| |X|^2 \|_{2p}^2 + \| |X|^2 \|_{1/2} \| |X|^2 \|_p \right) \left(\mathbb{E}|X|^\beta \right)^{1/q} \quad (44)$$

$$\leq \frac{64 \cdot 2^{\beta/q} c^2}{R^{\beta/q}} \| |X|^2 \|_{2p}^2 \| |X|^2 \|_{\beta/2}^{\beta/(2q)}. \quad (45)$$

Since $\mathbb{E}|X|^2 = d$, radial symmetry implies $\text{Cov}(X) = \mathbf{I}$. Therefore, by Lemma 1, we have

$$\mathbb{E} \left[|X - X_*|^2 |\Pi(X - X_*) (\rho(X) - \rho(X_*))|^2 \right] \geq 2(d - 1)I(X|G). \quad (46)$$

Continuing from above, we have proved that

$$J(X) - 2J(X - X_*) \geq \frac{d-1}{R^2} I(X|G) - \frac{32 \cdot 2^{\beta/q} c^2}{R^{2+\beta/q}} \| |X|^2 \|_{2p}^2 \| |X|^2 \|_{\beta/2}^{\beta/(2q)}. \quad (47)$$

For any $s > 0$, Taking $R = \left(\frac{b(2+s)}{2a} \right)^{1/s}$ yields the identity

$$\frac{a}{R^2} - \frac{b}{R^{2+s}} = \frac{1}{1+2/s} \left(\frac{2/s}{b(1+2/s)} \right)^{2/s} a^{1+2/s}. \quad (48)$$

So, putting $\varepsilon = 2q/\beta$, $p = 1 + 1/(2\varepsilon)$ and simplifying, we obtain

$$J(X) - 2J(X - X_*) \geq \frac{(\varepsilon/8)^\varepsilon}{(4(1+\varepsilon))^{1+\varepsilon}} \left(\frac{1}{c \| |X|^2 \|_{2p}} \right)^{2\varepsilon} \left(\frac{1}{\| |X|^2 \|_{q/\varepsilon}} \right) ((d-1)I(X|G))^{1+\varepsilon} \quad (49)$$

$$\geq \frac{(\varepsilon/8)^\varepsilon}{(8(1+\varepsilon))^{1+\varepsilon}} \left(\frac{1}{c \| |X|^2 \|_{2p}} \right)^{2\varepsilon} \left(\frac{1}{\| |X|^2 \|_{q/\varepsilon}} \right) (dI(X|G))^{1+\varepsilon} \quad (50)$$

$$= \frac{(\varepsilon/8)^\varepsilon}{c^{2\varepsilon} (8(1+\varepsilon))^{1+\varepsilon}} \left(\frac{1}{\| |X|^2 \|_{2+\varepsilon^{-1}}} \right)^{1+2\varepsilon} (dI(X|G))^{1+\varepsilon} \quad (51)$$

$$= \frac{(\varepsilon/8)^\varepsilon \| |X|^2 \|_1^{1+\varepsilon}}{c^{2\varepsilon} (8(1+\varepsilon))^{1+\varepsilon} \| |X|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}} I(X|G)^{1+\varepsilon}, \quad (52)$$

where we have made use of the crude bound $(d-1)/d \geq 1/2$ and substituted $d = \mathbb{E}|X|^2 = \| |X|^2 \|_1$. \square

3.2 Entropy Production under Rescaled Convolution

As one would expect, we may ‘integrate up’ in Theorem 1 to obtain an entropic version. A precise version of the result stated in Section 1 is given as follows:

Theorem 2. *Let X, X_* be IID radially symmetric random vectors on \mathbb{R}^d , $d \geq 2$, with c -regular density f . For any $\varepsilon > 0$*

$$h\left(\frac{1}{\sqrt{2}}(X + X_*)\right) - h(X) \geq C_\varepsilon(X) D(X)^{1+\varepsilon}, \quad (53)$$

where

$$C_\varepsilon(X) = \left(\frac{d\varepsilon}{1+(d+2)\varepsilon} \right)^{1+2\varepsilon} \frac{2^4 (d/100)^\varepsilon}{(2^8(1+\varepsilon)(1+2\varepsilon))^{1+\varepsilon}} \cdot \frac{\| |X|^2 \|_1}{c^{2\varepsilon} \| |X|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}}. \quad (54)$$

Remark 2. *Although the constant $C_\varepsilon(X)$ appears to grow favorably with dimension d , this dimension-dependent growth can cancel to give a bound that is effectively dimension-free. An illustrative example follows the proof.*

Proof. Similar to before, the inequality (53) is scale-invariant. Indeed, all relative entropy terms are invariant to scaling $t : X \mapsto tX$, and we also have $C_\varepsilon(tX) = C_\varepsilon(X)$ due to tX being (c/t^2) -regular if X is c -regular and homogeneity of the norms. Thus, we may assume without loss of generality

that X is normalized so that $\mathbb{E}|X|^2 = d$. Next, define $W = \frac{1}{\sqrt{2}}(X + X_*)$, and let X_t, W_t denote the Ornstein-Uhlenbeck evolutes of X and W , respectively. That is, for $t \geq 0$

$$X_t = e^{-t}X + (1 - e^{-2t})^{1/2}G, \quad W_t = e^{-t}W + (1 - e^{-2t})^{1/2}G. \quad (55)$$

By Proposition 2, X_t is $(5ce^{2t})$ -regular for all $t \geq 0$. Noting that $\mathbb{E}|X_t|^2 = \mathbb{E}|X|^2$, an application of Theorem 1 gives

$$I(X_t|G) - I(W_t|G) \geq \frac{(\varepsilon/8)^\varepsilon \| |X_t|^2 \|_1^{1+\varepsilon}}{(5c)^{2\varepsilon} (8(1+\varepsilon))^{1+\varepsilon} \| |X_t|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}} e^{-4\varepsilon t} I(X_t|G)^{1+\varepsilon} \quad (56)$$

$$\begin{aligned} &\geq \frac{(\varepsilon/8)^\varepsilon \| |X|^2 \|_1^{1+\varepsilon}}{(5c)^{2\varepsilon} (8(1+\varepsilon))^{1+\varepsilon} (2(\frac{1+(2+d)\varepsilon}{d\varepsilon}))^{1+2\varepsilon} \| |X|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}} e^{-4\varepsilon t} I(X_t|G)^{1+\varepsilon} \quad (57) \\ &= \left(\frac{d\varepsilon}{2 + (2d+4)\varepsilon} \right)^{1+2\varepsilon} \frac{(\varepsilon/8)^\varepsilon \| |X|^2 \|_1^{1+\varepsilon}}{(5c)^{2\varepsilon} (8(1+\varepsilon))^{1+\varepsilon} \| |X|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}} e^{-4\varepsilon t} I(X_t|G)^{1+\varepsilon}, \end{aligned}$$

where (57) holds since, for $p \geq 1$,

$$\| |X_t|^2 \|_p = (\mathbb{E}|X_t|^{2p})^{1/p} \leq 2 (\mathbb{E}(e^{-2t}|X|^2 + (1 - e^{-2t})|G|^2)^p)^{1/p} \quad (58)$$

$$= 2 \| e^{-2t}|X|^2 + (1 - e^{-2t})|G|^2 \|_p \quad (59)$$

$$\leq 2 \left(e^{-2t} \| |X|^2 \|_p + (1 - e^{-2t}) \| |G|^2 \|_p \right) \quad (60)$$

$$\leq 2(1 + \frac{p}{d}) \| |X|^2 \|_p. \quad (61)$$

The bound (61) uses the fact that $|G|^2$ is a chi-squared random variable with d degrees of freedom, and hence (using $\mathbb{E}|X|^2 = d$):

$$\| |G|^2 \|_p = \left(2^p \frac{\Gamma(p + \frac{d}{2})}{\Gamma(\frac{d}{2})} \right)^{1/p} = \mathbb{E}|X|^2 \left(\frac{\Gamma(p + \frac{d}{2})}{\Gamma(\frac{d}{2}) (\frac{d}{2})^p} \right)^{1/p} \quad (62)$$

$$\leq \mathbb{E}|X|^2 \left(1 + \frac{p}{d} \right) \quad (63)$$

$$\leq \| |X|^2 \|_p \left(1 + \frac{p}{d} \right). \quad (64)$$

Now, the claim will follow by integrating both sides. Indeed, by the classical de Bruijn identity, we have

$$\int_0^\infty (I(X_t|G) - I(W_t|G)) dt = D(X|G) - D(W|G) = h(\frac{1}{\sqrt{2}}(X + X_*)) - h(X). \quad (65)$$

By Jensen's inequality,

$$\int_0^\infty e^{-4\varepsilon t} I(X_t|G)^{1+\varepsilon} dt \geq \frac{1}{(4\varepsilon)^\varepsilon} \left(\int_0^\infty e^{-4\varepsilon t} I(X_t|G) dt \right)^{1+\varepsilon} \quad (66)$$

$$\geq \frac{1}{(4\varepsilon)^\varepsilon} \left(\int_0^\infty I(X_{t+2\varepsilon t}|G) dt \right)^{1+\varepsilon} \quad (67)$$

$$= \frac{1}{(4\varepsilon)^\varepsilon (1 + 2\varepsilon)^{1+\varepsilon}} D(X|G)^{1+\varepsilon}, \quad (68)$$

where we used the bound $I(X_{t+s}|G) \leq e^{-2s} I(X_t|G)$ due to exponential decay of information along the semigroup (e.g., [19]), a change of variables, and the identity $\int_0^\infty I(X_t|G)dt = D(X|G)$. Thus, we have proved

$$\begin{aligned} & h\left(\frac{1}{\sqrt{2}}(X + X_*)\right) - h(X) \\ & \geq \left(\frac{d\varepsilon}{2 + (2d+4)\varepsilon}\right)^{1+2\varepsilon} \frac{(\varepsilon/8)^\varepsilon \|X\|_1^{1+\varepsilon}}{(5c)^{2\varepsilon} (8(1+\varepsilon))^{1+\varepsilon} \|X\|_{2+1/\varepsilon}^{1+2\varepsilon}} \cdot \frac{1}{(4\varepsilon)^\varepsilon (1+2\varepsilon)^{1+\varepsilon}} D(X|G)^{1+\varepsilon} \end{aligned} \quad (69)$$

$$= \left(\frac{d\varepsilon}{1 + (d+2)\varepsilon}\right)^{1+2\varepsilon} \frac{2^4 (d/100)^\varepsilon}{(2^8(1+\varepsilon)(1+2\varepsilon))^{1+\varepsilon}} \cdot \frac{\|X\|_1^{1+\varepsilon}}{c^{2\varepsilon} \|X\|_{2+1/\varepsilon}^{1+2\varepsilon}} D(X|G)^{1+\varepsilon}. \quad (70)$$

□

Example 1 (Centered Gaussian Mixtures). Define f_i to be the density associated to the centered Gaussian distribution with covariance $\sigma_i^2 \mathbf{I}$. Let X be a random vector on \mathbb{R}^d with density $f = \sum_{i=1}^n p_i f_i$. For convenience, assume $\sigma_1^2 = \min_i \sigma_i^2$ and that the σ_i 's are normalized so that $\sum_{i=1}^n p_i \sigma_i^2 = 1$ ($\{p_1, \dots, p_n\}$ is a probability vector). Then, for d large,

$$h\left(\frac{1}{\sqrt{2}}(X + X_*)\right) - h(X) \gtrsim \frac{1}{2^{15}} \cdot \frac{\sum_{i=1}^n p_i (\sigma_i - 1)^2}{\sigma_1^2 \sum_{i=1}^n p_i (\sigma_i^2 / \sigma_1^2)^3} D(X). \quad (71)$$

It is easy to verify that f is $(1/\sigma_1^2)$ -regular. Moreover, $\mathbb{E}|X|^2 = d$, so using the bound (63), we have

$$\|X\|_{2+1/\varepsilon}^2 \leq \left(\sum_{i=1}^n p_i \sigma_i^{2(2+1/\varepsilon)}\right)^{1/(2+1/\varepsilon)} (d + 2 + 1/\varepsilon) \quad (72)$$

Now, by Talagrand's inequality, we may lower bound $D(X) \geq \frac{d}{2} W_2^2\left(\frac{1}{\sqrt{d}}|X|, \frac{1}{\sqrt{d}}|G|\right)$ as we did in the discussion following Theorem 1. Putting everything together and simplifying, we obtain:

$$\begin{aligned} & h\left(\frac{1}{\sqrt{2}}(X + X_*)\right) - h(X) \\ & \geq \left(\frac{d\varepsilon}{1 + (d+2)\varepsilon}\right)^{2(1+2\varepsilon)} \frac{2^4 (1/200)^\varepsilon}{(2^8(1+\varepsilon)(1+2\varepsilon))^{1+\varepsilon}} \cdot \frac{W_2^{2\varepsilon}\left(\frac{1}{\sqrt{d}}|X|, \frac{1}{\sqrt{d}}|G|\right)}{\sigma_1^2 \left(\sum_{i=1}^n p_i (\sigma_i^2 / \sigma_1^2)^{2+1/\varepsilon}\right)^\varepsilon} D(X). \end{aligned} \quad (73)$$

An easy consequence of the LLN is the limit

$$W_2^2\left(\frac{1}{\sqrt{d}}|X|, \frac{1}{\sqrt{d}}|G|\right) \rightarrow \sum_{i=1}^n p_i (\sigma_i - 1)^2 \quad \text{as } d \rightarrow \infty, \quad (74)$$

so the claim follows by putting $\varepsilon = 1$ and crudely bounding.

It is straightforward to remove the explicit requirement in Theorem 2 for c -regularity:

Corollary 1. Let X, X_* be IID radially symmetric random vectors on \mathbb{R}^d , $d \geq 2$, with finite Fisher information. For any $\varepsilon > 0$

$$h\left(\frac{1}{\sqrt{2}}(X + X_*)\right) - h(X) \geq \tilde{C}_\varepsilon(X) \frac{D(X)^{1+3\varepsilon}}{I(X)^{2\varepsilon}}, \quad (75)$$

where

$$\tilde{C}_\varepsilon(X) = \left(\frac{d\varepsilon}{1 + (d+2)\varepsilon} \right)^{2+4\varepsilon} \frac{2^{12} (d/100)^\varepsilon}{(2^{17}(1+\varepsilon)(1+2\varepsilon))^{1+\varepsilon}} \cdot \frac{\| |X|^2 \|_1}{\| |X|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}}. \quad (76)$$

Remark 3. Although the requirement of c -regularity is eliminated in (75), we see that the statement of (75) is effectively the same as (53), with c -regularity being replaced by another measure of regularity of X , i.e., the relative Fisher information $I(X)$.

Proof. Observe that inequality (75) is invariant to scaling X . Indeed, for $t > 0$ we have $t^{4\varepsilon} \tilde{C}_\varepsilon(tX) = \tilde{C}_\varepsilon(X)$ and $t^2 I(tX|G^{tX}) = I(X|G^X)$. Therefore,

$$\tilde{C}_\varepsilon(tX) \frac{1}{I(tX|G^{tX})^{2\varepsilon}} = \tilde{C}_\varepsilon(X) \frac{1}{I(X|G^X)^{2\varepsilon}}. \quad (77)$$

Since $D(tX|G^{tX}) = D(X|G^X)$ and $h(\frac{1}{\sqrt{2}}(tX + tX_*)) - h(tX) = h(\frac{1}{\sqrt{2}}(X + X_*)) - h(X)$, the claim follows. Thus, as before, we will assume without loss of generality that X is normalized so that $\mathbb{E}|X|^2 = d$.

Next, define $W = \frac{1}{\sqrt{2}}(X + X_*)$, and let X_t, W_t denote the Ornstein-Uhlenbeck evolutes of X and W , respectively. That is, for $t \geq 0$

$$X_t = e^{-t}X + (1 - e^{-2t})^{1/2}G, \quad W_t = e^{-t}W + (1 - e^{-2t})^{1/2}G. \quad (78)$$

By Proposition 1, X_t is $4(1 - e^{-2t})^{-1}$ -regular for all $t \geq 0$. Noting that $\mathbb{E}|X_t|^2 = \mathbb{E}|X|^2$, an application of Theorem 2 gives

$$\begin{aligned} h(W_t) - h(X_t) &\geq \left(\frac{d\varepsilon}{1 + (d+2)\varepsilon} \right)^{1+2\varepsilon} \frac{2^8 (d/100)^\varepsilon}{(2^{12}(1+\varepsilon)(1+2\varepsilon))^{1+\varepsilon}} \cdot \frac{(1 - e^{-2t})^{2\varepsilon} \| |X_t|^2 \|_1}{\| |X_t|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}} D(X_t|G)^{1+\varepsilon} \end{aligned} \quad (79)$$

$$\geq \left(\frac{d\varepsilon}{1 + (d+2)\varepsilon} \right)^{1+2\varepsilon} \frac{2^8 (d/100)^\varepsilon}{(2^{12}(1+\varepsilon)(1+2\varepsilon))^{1+\varepsilon}} \cdot \frac{(1 - e^{-2t})^{2\varepsilon} \| |X|^2 \|_1}{(2(\frac{1+(2+d)\varepsilon}{d\varepsilon}))^{1+2\varepsilon} \| |X|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}} D(X_t|G)^{1+\varepsilon} \quad (80)$$

$$= \left(\frac{d\varepsilon}{1 + (d+2)\varepsilon} \right)^{2+4\varepsilon} \frac{2^9 (d/100)^\varepsilon}{(2^{14}(1+\varepsilon)(1+2\varepsilon))^{1+\varepsilon}} \cdot \frac{(1 - e^{-2t})^{2\varepsilon} \| |X|^2 \|_1}{\| |X|^2 \|_{2+1/\varepsilon}^{1+2\varepsilon}} D(X_t|G)^{1+\varepsilon}, \quad (81)$$

where (80) follows by the same logic as (57). Now, by de Bruijn's identity and the convolution inequality for Fisher information, it follows that

$$\frac{d}{dt} (h(W_t) - h(X_t)) = J(W_t) - J(X_t) \leq 0. \quad (82)$$

Thus, $h(W) - h(X) \geq h(W_t) - h(X_t)$. The map $t \mapsto D(X_t|G)$ is continuous on $t \in [0, \infty)$ (e.g., [4]). Hence, using the fact that $\frac{d}{dt} D(X_t|G) = -I(X_t|G)$ (de Bruijn's identity) and that $t \mapsto I(X_t|G)$ is monotone-decreasing on $t \in [0, \infty)$ (convolution inequality for Fisher information), we have the inequality $D(X_t|G) \geq D(X|G) - tI(X|G)$.

Thus, to finish the proof, put $t = \frac{D(X|G)}{2I(X|G)}$ and note that $1 - e^{-x} \geq \frac{1}{\sqrt{2}}x$ for $x \in [0, 1/4]$, which applies for our choice of t due to the log Sobolev inequality, i.e., $\frac{1}{2}I(X|G) \geq D(X|G)$. \square

3.3 Beyond radial symmetry

Theorem 2 can be immediately extended to distributions that are radially symmetric, modulo an affine transformation. Indeed, in this case, we can apply the appropriate linear transformation $X \mapsto A(X - \mu)$, and then invoke the formulae for the behavior of entropy under such transformations to bound the desired entropy jump. A similar statement may be made for Theorem 1.

Less immediately, there is potential to bound the entropy jumps associated to a general non-symmetric random vector X by considering its symmetric decreasing rearrangement X^* (see, e.g., [20] for definition). In this case, it is known that $h(X) = h(X^*)$ and $h(X + X_*) \geq h(X^* + X_*)$ for X, X_* independent [20]. Thus, we might expect to be able to bound the entropy jump $h(\frac{1}{\sqrt{2}}(X + X_*)) - h(X)$ from below by $D(X^*)$. There are two issues here that must be considered in order to proceed along this route. The first issue is fundamental: we cannot expect to bound $D(X^*)$ by $D(X)$ in general (e.g., consider X to be a rearrangement of G). The second issue is technical and arises from the assumed regularity conditions. In particular, one would need to establish conditions under which the regularity of X^* can be appropriately controlled by the regularity of X (either in the sense of c -regularity or relative Fisher information).

In any case, it may be possible to apply the general proof idea to non-symmetric X . In particular, radial symmetry of X was only critically used in passing from (31) to (32). In general, by the projection property of the score function, we have

$$\begin{aligned} & \mathbb{E} |2\Pi(W)\rho_W(W) - \Pi(X - X_*)(\rho(X) - \rho(X_*))|^2 \\ &= \mathbb{E} |\Pi(X - X_*)(\rho(X) - \rho(X_*))|^2 - 4\mathbb{E} |\Pi(W)\rho_W(W)|^2, \end{aligned} \quad (83)$$

so it would be sufficient to bound

$$4\mathbb{E} |\Pi(W)\rho_W(W)|^2 < \alpha \mathbb{E} |\Pi(X - X_*)(\rho(X) - \rho(X_*))|^2 \quad (84)$$

for some $\alpha < 1$ (possibly depending on the distribution of X) in order for the proof to carry over to general X . Note that the distribution of $W = X - X_*$ is symmetric in general, but not radially symmetric, which may be useful toward establishing such a bound.

4 Estimate of the Deficit in the Log Sobolev Inequality

The problem of quantitatively bounding the deficit in the logarithmic Sobolev inequality has received considerable attention lately (e.g., [21–23]), but the problem of obtaining a satisfactory estimate that is dimension-free remains open in general. In this section, we apply our previous results to bound the deficit in the LSI for radially symmetric functions.

For a random vector X on \mathbb{R}^d , define the entropy power of X as

$$N(X) = \frac{1}{2\pi e} \exp\left(\frac{2}{d}h(X)\right). \quad (85)$$

The following entropy power inequality was proved by the author in [24]:

Theorem 3. *Let X, X_*, Z be independent random vectors on \mathbb{R}^d , with Z being Gaussian. Then*

$$N(X + X_* + Z)N(Z) + N(X)N(X_*) \leq N(X + Z)N(X_* + Z). \quad (86)$$

Particularizing to the case where $Z \sim N(0, I)$, we have the immediate corollary:

$$\frac{N(X + \sqrt{t}Z)N(X_* + \sqrt{t}Z) - N(X)N(X_*)}{t} \geq N(X + X_* + \sqrt{t}Z) \geq N(X + X_*). \quad (87)$$

Letting $t \rightarrow 0$ and applying de Bruin's identity gives the inequality:

$$dN(X + X_*) \leq N(X)N(X_*) (J(X) + J(X_*)). \quad (88)$$

Supposing X, X_* are identically distributed, we obtain the following improvement of Stam's inequality [3], which states that $N(X)J(X) \geq d$:

$$\frac{1}{d}N(X)J(X) \geq \exp \left\{ \frac{2}{d} \left(h\left(\frac{1}{\sqrt{2}}(X + X_*)\right) - h(X) \right) \right\}. \quad (89)$$

By work of Carlen [25], it is well-known that Stam's inequality is equivalent to Gross' log Sobolev inequality for the Gaussian measure [26]; i.e., $\frac{1}{2}I(X|G) \geq D(X|G)$. Using the inequality $\log x \leq x - 1$, it is straightforward to convert (89) into the following improved log Sobolev inequality:

$$\delta_{\text{LSI}}(X) \geq \left(h\left(\frac{1}{\sqrt{2}}(X + X_*)\right) - h(X) \right), \quad (90)$$

where $\delta_{\text{LSI}}(X) := \frac{1}{2}I(X|G) - D(X|G)$ denotes the deficit in the log Sobolev inequality.

In view of Corollary 1, we have established the following logarithmic Sobolev inequality in quantitative form for radially symmetric X :

Theorem 4. *Let X be a radially symmetric random vector on \mathbb{R}^d , $d \geq 2$, with $D(X) < \infty$. It holds that*

$$\delta_{\text{LSI}}(X) \geq \sup_{\varepsilon > 0} \tilde{C}_\varepsilon(X) \frac{D(X)^{1+3\varepsilon}}{I(X)^{2\varepsilon}}, \quad (91)$$

where $\tilde{C}_\varepsilon(X)$ is defined as in (76).

Theorem 4 yields a quantitative stability result for the LSI whenever $\mathbb{E}|X|^p < \infty$ for some $p > 4$. As an illustrative example, suppose X is normalized such that $\mathbb{E}|X|^2 = d$ (so that $D(X) = D(X|G)$ and $I(X) = I(X|G)$) and $\mathbb{E}|X|^8 < \infty$. Then, a clean bound is obtained by setting $\varepsilon = 1/2$, rearranging, solving a quadratic inequality and bounding constant terms:

$$\frac{1}{2}I(X) \geq D(X) \frac{1}{2} \left(1 + \sqrt{1 + \frac{1}{10^8} \frac{\| |X|^2 \|_1^{5/2}}{\| |X|^2 \|_4^2} \sqrt{D(X)}} \right). \quad (92)$$

Note that this is a very strong stability result. Indeed, it is a straightforward exercise (identify $\| |X|^2 \|_1 = d$, apply Minkowski's inequality to $\| |X|^2 \|_4$ and simplify) to show that

$$D(X) \leq 10^8 \max \left\{ \delta_{\text{LSI}}(X), \delta_{\text{LSI}}^{1/2}(X) \sqrt{\frac{\mathbb{E}|X_1|^8}{d}} \right\}. \quad (93)$$

We emphasize that $\mathbb{E}|X_1|^8$ is the eighth moment of the *one-dimensional* random variable X_1 , the first coordinate of X . Hence, we would generally expect that $\mathbb{E}|X_1|^8 \ll d$ in high dimension.

If X is known to be c -regular for some c , then we can establish the following stability estimate, which has no explicit dependence on dimension:

Theorem 5. *Let X be a radially symmetric random vector on \mathbb{R}^d , $d \geq 2$, with c -regular density f . For any $\varepsilon > 0$*

$$\delta_{\text{LSI}}(X) \geq \frac{1}{4} K_\varepsilon(X) I(X)^{1+\varepsilon}, \quad (94)$$

where $K_\varepsilon(X)$ is as defined in (19).

Proof. All quantities in (90) are invariant to translations of X , so we assume without loss of generality that $\mathbb{E}X = 0$ for the moment. In this case, we may write

$$\delta_{\text{LSI}}(X) \geq D(X|G) - D\left(\frac{1}{\sqrt{2}}(X + X_*)|G\right) \quad (95)$$

$$= -\delta_{\text{LSI}}(X) + \frac{1}{2} \left(I(X|G) - I\left(\frac{1}{\sqrt{2}}(X + X_*)|G\right) \right) + \delta_{\text{LSI}}\left(\frac{1}{\sqrt{2}}(X + X_*)\right) \quad (96)$$

$$\geq -\delta_{\text{LSI}}(X) + \frac{1}{2} \left(J(X) - J\left(\frac{1}{\sqrt{2}}(X + X_*)\right) \right). \quad (97)$$

Thus, in view of Theorem 1, we have proved the claim. \square

Remark 4. *Although Theorems 4 and 5 consider radially symmetric densities, we see from (90) and the following inequalities that $\delta_{\text{LSI}}(X)$ dominates entropy production and dissipation of Fisher information in general. Stated another way, any quantitative lower bound on entropy production (or Fisher information dissipation) will provide a lower bound on $\delta_{\text{LSI}}(X)$.*

In closing, we mention here that a quantitative form of the sharp Sobolev inequality was established by Cianchi, Fusco, Maggi and Pratelli [27] through a reduction of the general inequality to the setting of radially symmetric functions by considering spherically symmetric rearrangements. On this note, from [20, Corollary 8.7], it is easy to verify that $\delta_{\text{LSI}}(X) \geq \delta_{\text{LSI}}(X^*)$, where as before, X^* is the symmetric decreasing rearrangement of X . Provided X^* has regular density, we obtain

$$\delta_{\text{LSI}}(X) \geq \frac{1}{4} K_\varepsilon(X^*) I(X^*)^{1+\varepsilon}. \quad (98)$$

This should be compared to a result by Bobkov, Gozlan, Roberto and Samson [22] (see also [23]): If X is a random vector on \mathbb{R}^d with smooth density $p = e^{-V}$ satisfying $V'' \geq \varepsilon I$ for some $\varepsilon > 0$, then

$$\delta_{\text{LSI}}(X) \geq c_\varepsilon W_2^2(\bar{X}, G), \quad (99)$$

where c_ε is a constant depending on ε , only, and \bar{X} corresponds to a rearrangement of the vector X such that its one-dimensional marginals $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_d$ form a martingale [22]. We remark that the log-concavity assumption can be more restrictive than our regularity assumptions, but the symmetry of \bar{X} is less restrictive than radial symmetry. We also mention that Fathi, Indrei and Ledoux [21] have established another quantitative estimate on $\delta_{\text{LSI}}(X)$, under the assumption that X satisfies a Poincaré inequality. Namely, if X is a centered random vector that has spectral gap ε , then

$$\delta_{\text{LSI}}(X) \geq c_\varepsilon I(X), \quad (100)$$

where c_ε is a constant depending on ε , only. It is interesting to recall from our above discussion that stability of entropy jumps (in dimension one) has previously been established under a spectral gap condition [5]. We have now seen in (90) that the two stability problems are closely connected.

Appendix: Proof of Propositions 2 and 3

Proof of Proposition 2. First, suppose X is c -regular, and let $V = X + Z$, $Z \sim N(0, \sigma^2 I)$. We claim that V is $(5c)$ -regular. Toward this end, let f_X , f_Z and f_V denote the densities of X, Z, V , respectively. Now,

$$|\nabla f_V(v)| = \left| \int (\nabla f_X(v-z)) f_Z(z) dz \right| \quad (101)$$

$$\leq \int |\nabla f_X(v-z)| f_Z(z) dz \quad (102)$$

$$\leq c \int f_X(v-z) (|v-z| + \mathbb{E}|X|) f_Z(z) dz \quad (103)$$

$$\leq c f_V(v) (|v| + \mathbb{E}|X|) + c \int |z| f_X(v-z) f_Z(z) dz \quad (104)$$

$$= c f_V(v) (|v| + \mathbb{E}|X|) + c f_V(v) \mathbb{E}[|X-v| | V=v] \quad (105)$$

$$\leq c f_V(v) (|v| + \mathbb{E}|X|) + c f_V(v) (3|v| + 4\mathbb{E}|X|) \quad (106)$$

$$\leq 5c f_V(v) (|v| + \mathbb{E}|V|). \quad (107)$$

A proof of the only nontrivial inequality $\mathbb{E}[|X-v| | V=v] \leq 3|v| + 4\mathbb{E}|X|$ can be found in [18, Proposition 2]. Since V is smooth with nonvanishing density, $\nabla \log f_V = \frac{\nabla f_V}{f_V}$; this completes the proof of the claim.

By a change of variables, we observe that $e^{-t}X$ is (ce^{2t}) -regular. Combining with the previous claim finishes the proof. \square

Remark 5. *The claim of Proposition 2 may be strengthened. Indeed, if X is c -regular, then X_t is $(5c+4)$ -regular for all $t \geq 0$. To see this, note that Proposition 1 establishes that X_t is $(4/(1-e^{-2t}))$ -regular. However, Proposition 2 shows that X_t is $(5ce^{2t})$ -regular; maximizing the minimum of these two quantities over $t \geq 0$ establishes the strengthened claim. The weaker claim is more convenient for our purposes, so that regularity of X_t is a multiple of c . This ensures that the inequalities are invariant to scaling of X .*

Proof of Proposition 3. The proof is elementary, but provided here for completeness. Define $X_0 = \sqrt{d}R_0U$, where as before U is uniform on \mathbb{S}^{d-1} and independent of R_0 . Now, let

$$X = \sqrt{1-\varepsilon}X_0 + \sqrt{\varepsilon}G, \quad (108)$$

and set $R = \frac{1}{\sqrt{d}}|X|$. Clearly, $\mathbb{E}R^2 = 1$ and, by Proposition 1, X is $(4/\varepsilon)$ -regular. Now, observe that

$$F_R(r) = \Pr \left\{ \frac{1}{\sqrt{d}} |\sqrt{1-\varepsilon}X_0 + \sqrt{\varepsilon}G| \leq r \right\} \quad (109)$$

$$\geq \Pr \left\{ \frac{\sqrt{1-\varepsilon}}{\sqrt{d}} |X_0| + \frac{\sqrt{\varepsilon}}{\sqrt{d}} |G| \leq r \right\} \quad (110)$$

$$\geq \Pr \left\{ \frac{\sqrt{1-\varepsilon}}{\sqrt{d}} |X_0| \leq r - \sqrt{(t+1)\varepsilon} \right\} \Pr \left\{ \frac{\sqrt{\varepsilon}}{\sqrt{d}} |G| \leq \sqrt{(t+1)\varepsilon} \right\} \quad (111)$$

$$= F_{R_0} \left(\frac{r - \sqrt{(t+1)\varepsilon}}{\sqrt{1-\varepsilon}} \right) (1 - \Pr \left\{ \frac{1}{d} |G|^2 - 1 > t \right\}) \quad (112)$$

$$\geq F_{R_0} \left(\frac{r - \sqrt{(t+1)\varepsilon}}{\sqrt{1-\varepsilon}} \right) - e^{-dt^2/8}, \quad (113)$$

where the final inequality is due to the tail bound for a chi-squared random variable with d degrees of freedom. The other direction is similar:

$$F_R(r) = \Pr \left\{ \frac{1}{\sqrt{d}} |\sqrt{1-\varepsilon} X_0 + \sqrt{\varepsilon} G| \leq r \right\} \quad (114)$$

$$\leq \Pr \left\{ \frac{\sqrt{1-\varepsilon}}{\sqrt{d}} |X_0| - \frac{\sqrt{\varepsilon}}{\sqrt{d}} |G| \leq r \right\} \quad (115)$$

$$\leq \Pr \left\{ \frac{\sqrt{1-\varepsilon}}{\sqrt{d}} |X_0| \leq r + \sqrt{(t+1)\varepsilon} \right\} + \Pr \left\{ \frac{\sqrt{\varepsilon}}{\sqrt{d}} |G| > \sqrt{(t+1)\varepsilon} \right\} \quad (116)$$

$$= F_{R_0} \left(\frac{r + \sqrt{(t+1)\varepsilon}}{\sqrt{1-\varepsilon}} \right) + e^{-dt^2/8}. \quad (117)$$

□

Acknowledgment

This work was supported in part by NSF grants CCF-1528132 and CCF-0939370 (Center for Science of Information).

References

- [1] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:623–656, 1948.
- [2] Nelson M Blachman. The convolution inequality for entropy powers. *Information Theory, IEEE Transactions on*, 11(2):267–271, 1965.
- [3] A.J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, 1959.
- [4] E. A. Carlen and A. Soffer. Entropy production by block variable summation and central limit theorems. *Communications in mathematical physics*, 140(2):339–371, 1991.
- [5] Keith Ball, Franck Barthe, and Assaf Naor. Entropy jumps in the presence of a spectral gap. *Duke Mathematical Journal*, 119(1):41–63, 2003.
- [6] Oliver Johnson and Andrew Barron. Fisher information inequalities and the central limit theorem. *Probability Theory and Related Fields*, 129(3):391–409, 2004.
- [7] Oliver Johnson. *Information theory and the central limit theorem*, volume 8. World Scientific, 2004.
- [8] Keith Ball and Van Hoang Nguyen. Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Mathematica*, 213(1):81–96, 2012.
- [9] Giuseppe Toscani. A strengthened entropy power inequality for log-concave densities. *IEEE Transactions on Information Theory*, 61(12):6550–6559, 2015.
- [10] Sergey G Bobkov, Gennadiy P Chistyakov, and Friedrich Götze. Rate of convergence and edgeworth-type expansion in the entropic central limit theorem. *The Annals of Probability*, 41(4):2479–2512, 2013.

- [11] Sergey G Bobkov, Gennadiy P Chistyakov, and Friedrich Götze. Berry-esseen bounds in the entropic central limit theorem. *Probability Theory and Related Fields*, 159(3-4):435, 2014.
- [12] Michel Ledoux, Ivan Nourdin, and Giovanni Peccati. A Stein deficit for the logarithmic Sobolev inequality. *arXiv preprint arXiv:1602.08235*, 2016.
- [13] Cédric Villani. A review of mathematical topics in collisional kinetic theory. *Handbook of mathematical fluid dynamics*, 1:71–305, 2002.
- [14] E. A. Carlen and M. C. Carvalho. Strict entropy production bounds and stability of the rate of convergence to equilibrium for the boltzmann equation. *Journal of statistical physics*, 67(3-4):575–608, 1992.
- [15] G. Toscani and C. Villani. Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. *Communications in mathematical physics*, 203(3):667–706, 1999.
- [16] Cédric Villani. Cercignani’s conjecture is sometimes true and always almost true. *Communications in mathematical physics*, 234(3):455–490, 2003.
- [17] Laurent Desvillettes and Cédric Villani. On the spatially homogeneous Landau equation for hard potentials part II: H-theorem and applications. *Communications in Partial Differential Equations*, 25(1-2):261–298, 2000.
- [18] Y. Polyanskiy and Y. Wu. Wasserstein continuity of entropy and outer bounds for interference channels. *IEEE Transactions on Information Theory*, 62(7):3992–4002, July 2016.
- [19] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer Science & Business Media, 2013.
- [20] Liyao Wang and Mokshay Madiman. Beyond the entropy power inequality, via rearrangements. *Information Theory, IEEE Transactions on*, 60(9):5116–5137, 2014.
- [21] Max Fathi, Emanuel Indrei, and Michel Ledoux. Quantitative logarithmic Sobolev inequalities and stability estimates. *arXiv preprint arXiv:1410.6922*, 2014.
- [22] Sergey G Bobkov, Nathael Gozlan, Cyril Roberto, and P-M Samson. Bounds on the deficit in the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 267(11):4110–4138, 2014.
- [23] Emanuel Indrei and Diego Marcon. A quantitative log-Sobolev inequality for a two parameter family of functions. *International Mathematics Research Notices*, page rnt138, 2013.
- [24] Thomas A Courtade. Strengthening the entropy power inequality. *arXiv preprint arXiv:1602.03033*, 2016.
- [25] Eric A Carlen. Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 101(1):194–211, 1991.
- [26] Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

- [27] Andrea Cianchi, Nicola Fusco, Francesco Maggi, and Aldo Pratelli. The sharp sobolev inequality in quantitative form. *Journal of the European Mathematical Society*, 11(5):1105–1139, 2009.